

学校编码: 10384

分类号_____密级_____

学号: X2006230075

UDC_____

廈門大學

碩 士 學 位 論 文

基于 CWM 的商业银行元数据仓库的
研究与应用

Research and Implementation of Commercial Bank
Metadata Warehouse Based on CWM

謝 澤 添

指导教师姓名: 王备战 副教授

专 业 名 称: 软 件 工 程

论文提交日期: 2008 年 8 月

论文答辩时间: 2008 年 9 月

学位授予日期: 年 月

答辩委员会主席: _____

评 阅 人: _____

2008 年 月

厦门大学博硕士论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士论文摘要库

摘 要

数据仓库技术是在数据库基础上发展而来的新一代信息管理技术,主要用于支持企业信息集成、数据挖掘、企业决策支持等应用。在数据仓库建设过程中,由于各工具厂商采用不同的元数据标准,使得依靠这些工具进行数据集成、数据共享显得十分困难,由此引发了人们对元数据的研究。2001 年 OMG 组织在其已制定的规范 UML、MOF、XMI 的基础上提出公共仓库元模型(CWM)。

本文首先系统地介绍了 CWM 的体系结构,对 CWM 中每个包的内容以及包与包之间的关系都进行了深入的剖析。在对 CWM 规范、对历史上元数据的管理策略和元数据解决方案的深入研究,同时对该商业银行各类信息系统的集成需求分析之后,论文决定采纳元数据仓库(MDW)的管理策略,在该商业银行搭建基于 CWM 的元数据仓库的集成体系结构。

接着本文从整个元数据生命周期出发,有特色地设计了元数据仓库的六个组件:来源层、集中层、仓储、集市、管理层、交换层,全面管理各生命周期的元数据。对于元数据仓库的核心组件——元数据仓储,论文以较大篇幅详述了其三层架构的设计,同时论文进行了大胆而有效的实践,这种三层架构具备高度的适应性、扩展性,不仅在国内处于领先地位,开发的产品也可与国际知名的元数据产品相提并论。与传统的数据仓库 DB/DW 两层结构不同,论文所在的商业银行正积极构建包含数据仓库在内的三层数据架构,即 DB/ODS/DW 三层结构的信息体系,因此论文是站在企业的角度来考虑元数据的,相较以往基于数据仓库的元数据研究,可以说是又前进了一大步。

最后论文以实践来说明,如何结合 CWM 规范在元数据仓储层设计数据字典、数据转换、数据质量、业务应用四大主题的元模型,并在元数据仓库中对四大主题的元数据进行集成与管理,有效地支持了该商业银行信息体系建设的工作。

关键词: 公共仓库元模型; 元数据仓库; 元-元模型

厦门大学博硕士论文摘要库

Abstract

Data Warehouse (DW) is the new information management technology evolved from database. It is prevailing in Enterprise Information Integration, Data Mining and Enterprise Decision Making and so on. The metadata standard of software development tools that are used for building the DW is different among these tools. Therefore the tools aren't able to collaborate with each other effectively. They have to be involved in the complicated mapping between metadata. In order to unify metadata management strategies of the vendors of DW and smooth metadata interchange, Object Management Organization (OMG) brought forward the Common Warehouse Metamodel (CWM) based the specification of UML, MOF and XMI.

Firstly this dissertation in a systematic way presents an introduction of the structure of CWM. The packages and the associations among packages are analyzed in detail. With the comprehensive analysis of the requirement of almost all the information system and the study of metadata solutions in history, a Metadata Warehouse strategy based on CWM is adopted.

Secondly, a CWM based Metadata Warehouse (MDW) is designed and implemented on the viewpoint of the whole life-cycle of metadata. MDW constitutes of six components, namely source layer, integration layer, repository, marts, management layer and delivery layer. This dissertation details the design and implementation of such six components especially for the metadata repository. Metadata repository is the core component of MDW and built on the basis of three-tier architecture consisting of M1, M2 and M3. Such three-tier architecture has both a high expansibility and adaptability. It is in leading place at home in the field of study and implementation of Metadata and can be mentioned with the big tools of Metadata outside China in the future.

Finally, how to design the meta-model and manage the metadata of the four domains of data dictionary, data transformation, data quality and business application based on CWM is described. The status and insufficiency is also indicated in this dissertation.

The commercial bank the author working for is focusing on three-tier DB/ODS/DW architecture. It is different from the traditional two-tier DB/DW architecture. The author hopes to seek the solutions beyond DW. This research and implementation make the Metadata stride forward one step greatly.

Key Words: CWM; Metadata Warehouse; Meta-Meta Model

厦门大学博硕士论文摘要库

目 录

第一章 绪 论	1
1.1 研究背景	1
1.2 国内外研究现状	2
1.3 论文主要内容及特色	3
1.4 论文组织结构	4
第二章 元数据与公共仓库元模型	6
2.1 元数据概论	6
2.1.1 元数据的分类	7
2.1.2 元数据的作用	8
2.1.3 元数据面临的挑战	9
2.2 公共仓库元模型	10
2.2.1 对象模型层	11
2.2.2 基础层	17
2.2.3 资源层	26
2.2.4 分析层	34
2.2.5 管理层	41
2.3 小结	44
第三章 元数据的管理策略和集成体系结构	46
3.1 元数据管理策略	46
3.1.1 搭建元数据交换的途径	47
3.1.2 搭建元数据存储库	47
3.1.3 搭建元数据仓库	48
3.2 元数据集成体系结构	48
3.2.1 点到点的元数据体系结构	48
3.2.2 中央存储库的元数据集成体系结构	49
3.2.3 基于模型的元数据集成体系结构	50
3.3 基于 CWM 的元数据仓库的集成体系结构	51
3.4 小结	52
第四章 元数据仓库的设计与实现	53
4.1 元数据来源层	53
4.2 元数据集中层	54
4.3 元数据仓储	55
4.3.1 仓储的设计	55
4.3.2 仓储的实现	56
4.4 元数据集市	62
4.5 元数据管理层	62
4.5.1 元数据管理层的设计要点	63
4.5.2 元数据管理层的职责	64

4.6 元数据交换层.....	65
4.6.1 交换的基本原则.....	66
4.6.2 交换方式.....	67
4.7 小结	68
第五章 元数据仓库的应用	69
5.1 数据字典	69
5.2 数据转换	70
5.3 数据质量	71
5.3.1 数据质量元数据.....	71
5.3.2 数据质量的全局分析.....	72
5.4 业务应用	73
5.5 小结	74
第六章 总结与展望	75
6.1 总结	75
6.2 展望	76
参考文献	78
攻读硕士期间科研成果	81
致 谢	82

Contents

CHAPTER 1 INTRODUCTION	1
1.1 BACKGROUND AND SIGNIFICANCE	1
1.2 RESEARCH OF METADATA	2
1.3 MAIN CONTENTS AND FEATURES.....	3
1.4 TEXT STRUCTURE	4
CHAPTER 2 METADATA AND CWM	6
2.1 CONSPECTUS OF METADATA.....	6
2.1.1 Kinds of Metadata	7
2.1.2 Role of Metadata	8
2.1.3 Challenge of Metadata	9
2.2 CWM INTRODUCTION	10
2.2.1 Object Model Layer	11
2.2.2 Foundation Layer	17
2.2.3 Resource Layer.....	26
2.2.4 Analysis Layer	34
2.2.5 Management Layer.....	41
2.3 SECTION CONCLUSION.....	44
CHAPTER 3 STRATEGY AND INTEGRATED ARCHITECTURE OF METADATA	46
3.1 STRATEGY OF METADATA MANAGEMENT.....	46
3.1.1 Putting Up Metadata Interchange	46
3.1.2 Putting Up Metadata Repository.....	47
3.1.3 Putting Up Metadata Warehouse.....	48
3.2 INTEGRATED ARCHITECTURE OF METADATA.....	48
3.2.1 Point to Poing Architecture of Metadata	48
3.2.2 Repository of Integrated Architecture of Metadata	49
3.2.3 Integrated Architecture of Metadata Based on Model.....	50
3.3 META WAREHOUSE BASED ON CWM	51
3.4 SECTION CONCLUSION	52
CHAPTER 4 DESIGN AND IMPLEMENTATION FOR METADATA WAREHOUSE.....	53
4.1 METADATA SOURCING LAYER	53
4.2 METADATA INTEGRATION LAYER	54
4.3 METADATA REPOSITORY.....	55
4.3.1 Design for Repository	55
4.3.2 Implementation for Repository	56
4.4 METADATA MARTS	62
4.5 METADATA MANAGEMENT LAYER	62
4.5.1 Points of Design for Metadata Management Layer.....	63
4.5.2 Functions of Metadata Management Layer.....	64
4.6 METADATA INTERCHANGE LAYER.....	65
4.6.1 Basic Principles for Interchange of Metadata.....	66
4.6.2 Ways of Interchange of Metadata.....	67

4.7 SECTION CONCLUSION	68
CHAPTER 5 APPLICATION OF METADATA WAREHOUSE.....	69
5.1 DATA DICTIONARY	69
5.2 DATA TRANSFORMATION.....	70
5.3 DATA QUALITY	71
5.3.1 Data Quality Metadata	71
5.3.2 Global Analysis for Data Quality.....	72
5.4 BUSINESS APPLICATION	73
5.5 SECTION CONCLUSION	74
CHAPTER 6 CONCLUSION AND FUTURE	75
6.1 CONCLUSION	75
6.2 FUTURE	76
REFERENCES.....	78
RESEARCHS	81
ACKNOWLEDGEMENT.....	82

第一章 绪 论

1.1 研究背景

本世纪以来，全球经济和市场环境持续发生巨大而深刻的变化，国内商业银行所面临的竞争与挑战日益激烈。为能有效利用有限的资源，更好地服务于客户，持续创造利润，各商业银行纷纷加快了信息化进程，希望从其所积累的大量业务数据中挖掘出有价值的信息，支持商业银行决策。在此情况下，出现了数据仓库(Data Warehouse, 简称为DW)系统，从逻辑上、物理上分布的多个数据库(Database, 简称DB)系统中抽取数据、集成信息，并将集成后的信息存储在一个大规模的数据仓库中，从而来支持OALP(On-Line Analytical Processing, 联机分析处理)、数据挖掘(DataMining)、企业决策。

考虑到国内商业银行数据量的巨大及用户分析需求的不同层次，商业银行在建设DW的同时，也加快了CIF客户信息整合、CRM客户关系管理、ERP企业资源管理等信息项目建设的步伐，这种信息处理的多层次要求导致了一种新的数据环境——操作数据存储(Operational Data Store, 简称ODS)的建立。一方面它包含细节的、当前或接近当前的数据，可进行联机操作型处理；另一方面，它又是一种面向主题、集成的数据环境，可以辅助企业完成日常决策。与此同时ODS建立起底层结构来横贯整个企业的异构系统、应用、数据源等，完成业务操作系统和DW、CIF、CRM、ERP，以及其他重要的内部系统之间无缝地共享和交换数据的需要，这就在商业银行内部形成DB/ODS/DW三层结构的信息体系^[1,2,3]，具体如图1.1所示。

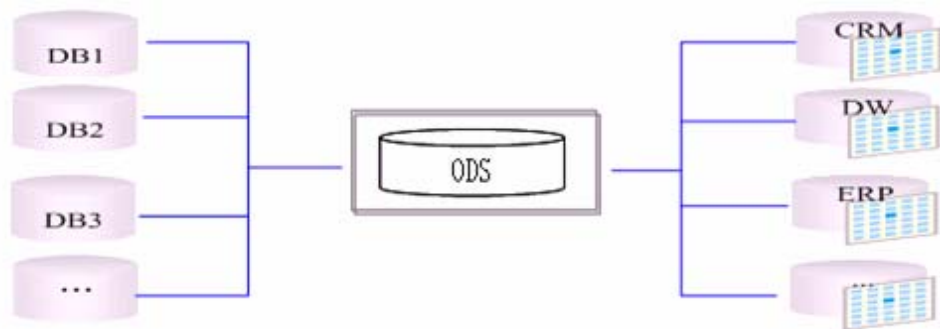


图 1.1 DB/ODS/DW 三层信息体系结构

资料来源：参考《基于 DB/ODS/DW 的 CRM 动态数据仓库》^[2]

当前,各商业银行的信息化建设已达到一定的水平,各业务领域广泛采用信息的计算机化处理。随着信息化建设的发展,信息共享日益重要。然而,国内商业银行多年“自发”式的信息化建设,缺乏“自觉”的IT系统总体规划,系统应用环境极其复杂,数据库结构和功能各异,所采用的建模、OLAP、ETL工具互不相同,这就造成商业银行内部IT系统各自为政,无法互通互联,形成“信息孤岛”^[4,5],这与商业银行日益增长的信息共享需求相背。

必须寻找一条能够将各种数据库产品、OLAP、ETL等工具有机结合在一起的途径,使其在建设数据仓库系统的过程中能够相互配合,在“信息孤岛”间架起沟通的桥梁,从而实现不同系统间的互联互通。而元数据(MetaData)管理正是解决这一问题的科学、有效的途径。

1.2 国内外研究现状

由于元数据的重要性,元数据的研究已引起国内外的广泛关注。OMG(Object Management Group, 对象管理组织)采纳了CWM^[6](Common Warehouse Metamodel, 公共仓库元模型)作为一个用于元数据交换的开放的业界标准,在数据仓库和业务分析领域为元数据定义公共的元模型^[6,7]以及基于XMI^[7](XML Metadata Interchange, XML元数据交换)的交换格式。目前,CWM已被IBM、UNISYS、NCR、Hyperion、Oracle等多家公司支持并打算整合到它们的下一代数据仓库与业务分析产品中。2000年的奥兰多国际会议上,IBM等六家公司展示了使用CWM在10种不同厂商的数据仓库产品间交换元数据^[8],交换过程代价小、灵活方便,这次展示可以说是数据仓库元数据交换的重大突破,2001、2002年分别在希尔顿阿纳海姆和圣安东尼奥召开的元数据年会上又进一步完善了这一过程,推广了CWM在各大厂商的使用^[9]。

2002年以来,我国元数据研究与应用一直呈现上升趋势,但多数基于图书情报领域^[10],而对于数据集成和业务分析领域的元数据管理技术的研究还不是很广泛、很深入,更何况最终是要把理论运用到实际中去。因此国内企业在建设元数据时,常常借助国内外市场上的元数据管理工具来实现。如NCR的MDS、DAG公司的MetaCenter、IBM的Metastage、Informatic的Superglue、ASG公司的Rochade,国内石竹公司的MetaOne。经过对这些工具的研究与试用,我们认为,这些工具支持开放的CWM标准,能进行多种工具的元数据抽取,并进行

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库